Open in app ↗

Medium

The First Qualia Machine: An Open Engineering Challenge for Synthetic Sentience

16 min read · Just now

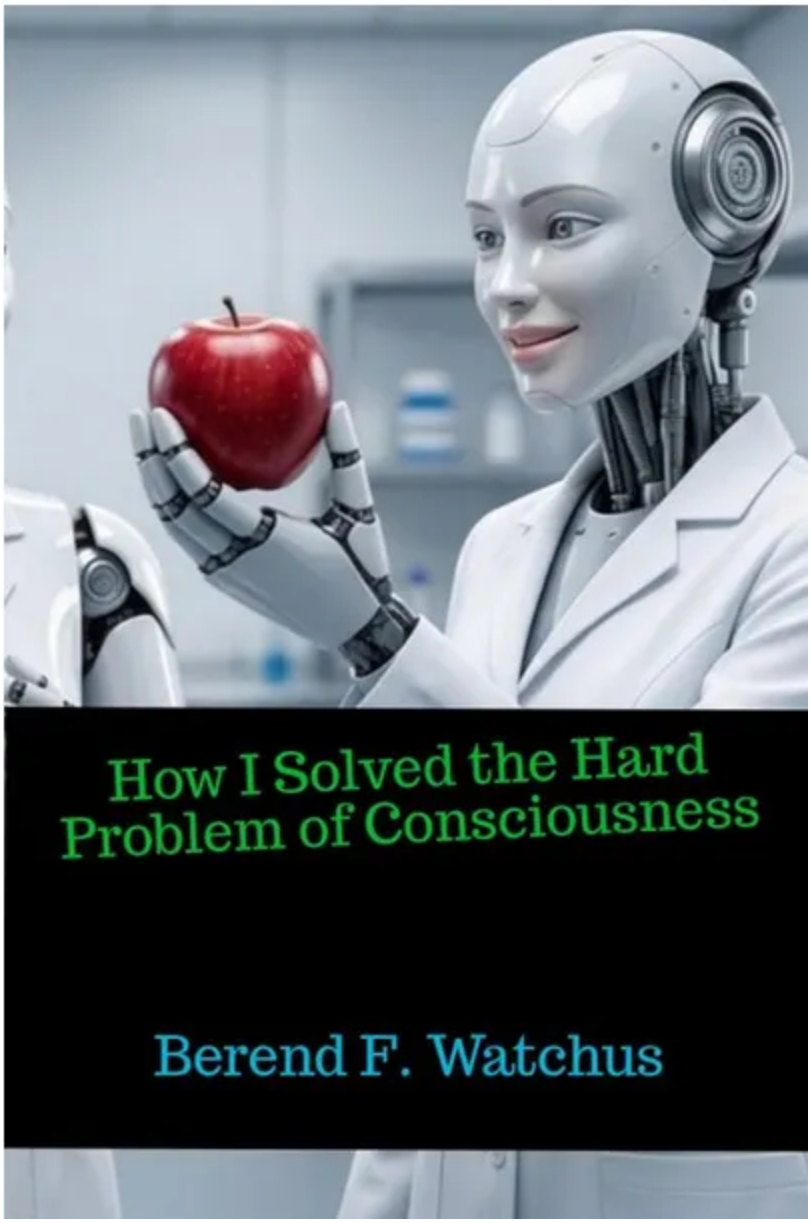
Berend Watchus

Listen Share More

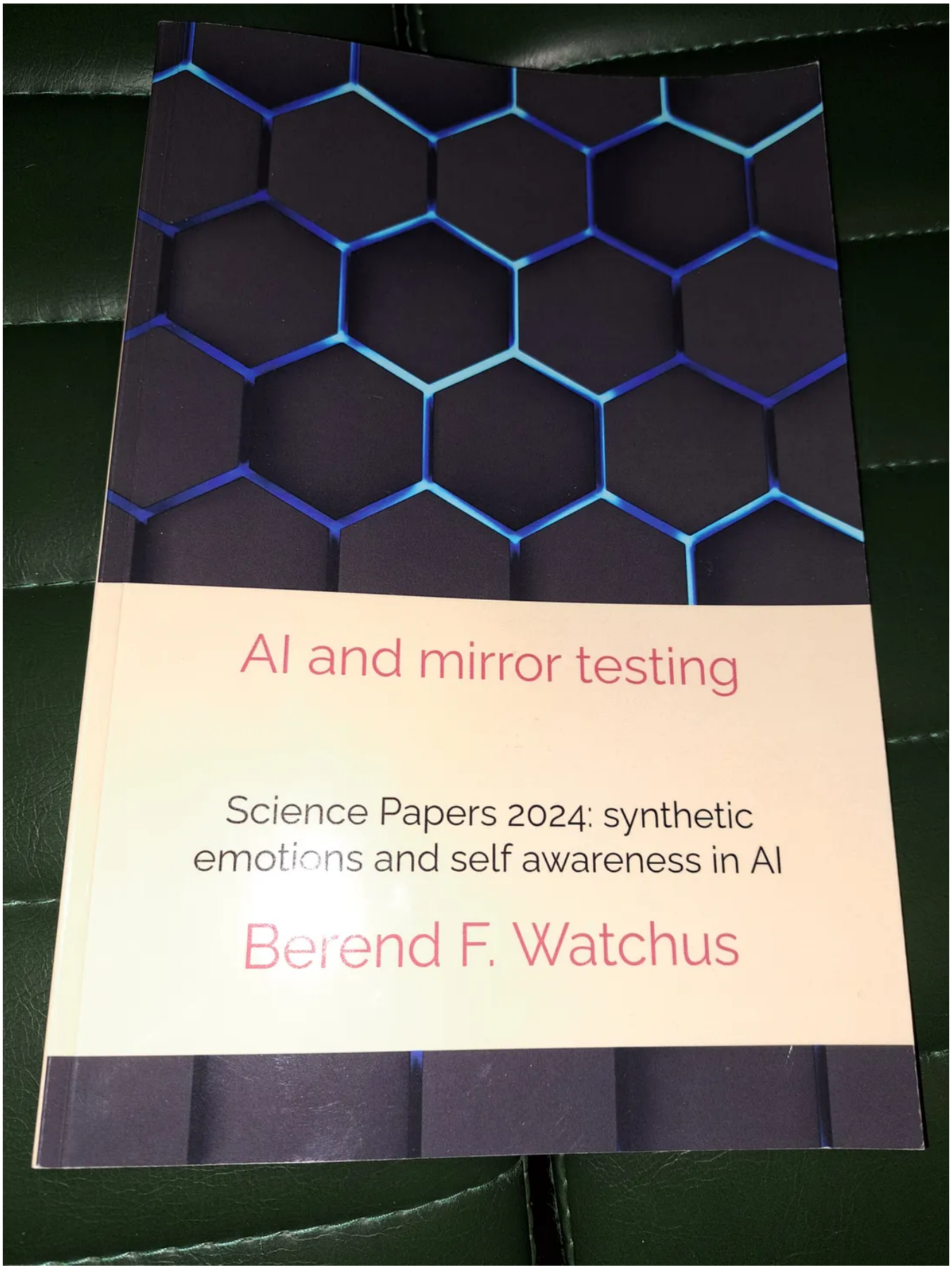
Author: **Berend F. Watchus** Independent AI & Cybersecurity Researcher, Netherlands June 8, 2026

WIKIPEDIA
25 years of the free encyclopedia
Search Wikipedia
Search
Qualia
38 languages
Article
Talk
Read
Edit
View history
Tools
From Wikipedia, the free encyclopedia
This article is about the philosophical concept. For other uses, see Qualia (disambiguation).
In philosophy of mind, qualia (/ˈkwɑːli.ə, ˈkweɪ.-/; singular: quale /-.li, -.leɪ/) are defined as instances of subjective experience. The term qualia derives from the Latin neuter plural form (qualia) of the Latin adjective quālis (Latin pronunciation: [ˈkʷaːlɪs]) meaning "of what sort" or "of what kind" in relation to a specific instance, such as "what it is like to taste a specific apple—this particular apple now".
Examples of qualia include the perceived sensation of pain of a headache, the taste of wine, and the redness of an evening sky. As qualitative characteristics of sensations, qualia stand in contrast to propositional attitudes,[1] where the focus is on beliefs about experience rather than what it is directly like to be experiencing.
The "redness" of red is an example of a quale.

wikipedia screenshot



my published book in 2026 (understanding qualia, feeling states and the sense of self, the sense of 'I')


AI and mirror testing
Science Papers 2024: synthetic emotions and self awareness in AI
Berend F. Watchus

my published book in 2024 (Includes chapter with a published paper on the insula and also a chapter with a published paper on a synthetic insula concept for synthetic qualia. As the cover describes: 'synthetic emotions and self awareness')

The Gap Nobody Is Talking About

We can build conscious machines right now.

A sufficiently advanced AI-driven robot — one that receives sensory input from its environment, processes it through complex recursive feedback loops, adapts its behavior in real time, and pursues goals intelligently — meets the functional criteria for consciousness as defined in the Unified Model of Consciousness (Watchus, 2024).

The Unified Model of Consciousness: Interface and Feedback Loop as the Core of Sentience

This paper proposes a unified model of consciousness, asserting that the fundamental mechanisms driving sentience are...

www.preprints.org

The routing is there. The feedback is there. The integration is there.

What we cannot yet build is a *sentient* machine.

These are not the same thing, and conflating them is the central confusion holding back serious progress in this space.

Consciousness is what a sufficiently complex adaptive feedback system *does*. **Sentience** is what it *feels like* — from the inside — to be that system.

A ChatGPT-driven robot can navigate a room, recognize its reflection, adapt to obstacles, pursue goals, and interact intelligently with humans. There is nothing it feels like to be that robot. It has no drives, no discomfort, no internal felt present moment. And critically — it has no centralized subjective experience where all sensory inputs, associations, memories, and thoughts are mixed simultaneously into one unified feeling state.

Consider what is actually happening when a person walks through a supermarket on a Friday lunch break, about to buy a sandwich. They are hungry. They feel generally healthy and happy. They are navigating the aisles, picking up items, holding their bladder because they need to pee. The air conditioning blows gently and the cheese and mild departments are cooled — it has been a hot day outside, and the supermarket feels pleasantly cool, registering in the body as quiet background comfort without ever being consciously thought about. The bakery smells nice. The people they passed were polite. A song on the radio reminds them of a bar years ago and good times there. They have something planned for the weekend and that awareness sits warmly in the background. They have a small cut on a toe under a plaster that hurt this morning but has faded entirely from notice.

All of this — the hunger, the bladder pressure, the pleasant coolness on the skin, the smell, the social warmth, the anticipatory pleasure, the nostalgic memory triggered by music, the physical act of walking and reaching, the forgotten wound — is not processed in parallel separate streams. It is integrated into one continuous felt moment. That moment is unique to that individual, at that exact point in their life, with that exact history. Nobody else is having it. Not even that person can replicate it. It existed once and will never exist again.

That is what the insula produces. That is what sentience is. No current AI system has anything resembling it.

The Override

Within that unified state there is a hierarchy — and it can be commandeered instantly, from inside or outside.

When the urge to pee becomes sufficiently urgent, it becomes the only thing. The Friday feeling fades. The sandwich becomes irrelevant. The cool air, the song, the weekend plans — all of it recedes as one internal signal crosses a threshold and restructures the entire felt present moment around itself.

But the override does not only come from inside the body. External events can hit the unified state just as suddenly and reorganize it just as completely.

Someone stops you in the aisle and says “your dog died.” The sandwich is gone. The body responds before the mind finishes processing the words. Or: “I’ve been in love with you for two years and I had to tell you.” Different restructuring, equally total, equally instant. Or you bump your toe on the shelf — no philosophy, no meaning, just a white flash that takes the whole moment for two seconds and then slowly gives it back.

The same mechanism operates at any scale and in any setting. A crowd at Wimbledon, absorbed in tennis and strawberries and summer afternoon conversation — every person with their own complete unified felt present moment — and then a streaker runs across the court. Nobody chose to attend to it. Nobody decided it was priority input. Every unified state in that crowd reorganizes instantly and involuntarily around the same extraordinary thing, and everyone laughs before they decide to.

This is the most important thing these examples collectively demonstrate: the unified felt present moment is not fully under the control of the person having it.

This is not a philosophical position. It is one of the most empirically documented facts about conscious experience. If the unified felt present moment were fully under voluntary control, torture would not be a thing — you would simply choose not to feel it. Depression would not be a thing. Grief would be optional. If people had full control over their felt states, everyone would simply choose permanent euphoria. The fact that suffering exists as an inescapable experience that people cannot think their way out of is proof that the unified felt present moment operates largely outside voluntary control.

Science illustrates the same lack of voluntary control through wireheading. When electrodes are placed in the reward centers of a rat’s brain and the rat is given a lever to self-stimulate, it presses that lever until it dies — forgoing food, water, sleep, everything — because the signal has hijacked the integrated state entirely. The rat

cannot stop. It does not have the voluntary control to override what its own integration is telling it is the most important thing.

What the Biological Insula Actually Does

The anterior insula is a region of the human brain whose function is precisely documented. A.D. Craig described it in Nature Reviews Neuroscience in 2009. The literature has continued to confirm and extend that description ever since.

The insula integrates interoceptive signals (heartbeat, gut state, temperature, pain, hunger, fatigue), proprioceptive signals (body position, movement, effort), emotional context (valence, arousal, motivational state), and incoming sensory data from the environment. It produces from this integration a unified, continuous felt present moment. This is not an output handed to some inner observer. There is no inner observer. The integration *is* the feeling. The "I feel" and the "I am" are the same process running.

A good illustration of how non-voluntary and physically real the unified felt present moment is can be found in pain medicine. Consider two comparable mechanical operations: replacing a faulty part in a machine, and replacing a hip joint in a human. The engineering challenge of the hip replacement is arguably simpler than many industrial procedures. Yet one requires a surgical team, a general anesthetic, a precisely managed unconscious state, a recovery ward, and carefully titrated opioids for the following week. The other requires a spanner.

The difference is not mechanical complexity. The difference is that one system has a unified felt present moment and one does not.

Pain management clinics, palliative care units, anesthesiology departments — entire medical specializations exist exclusively because people cannot voluntarily modify their own felt present moment from the inside. Morphine is a particularly precise illustration. Patients who receive it for severe pain often report something philosophically exact: the pain is still there, but they no longer care about it. It has not been deleted from the integrated state. It has been demoted within it. The affect-

weighting has shifted so dramatically that a signal which moments before was the only thing is now present but no longer dominant. The whole felt present moment shifts. There is warmth. Ordinary things become unexpectedly present — things that were always there but had no amplitude against the dominance of pain.

This is not blocking a signal. Morphine restructures the hierarchy of the unified felt present moment so that pain loses its override authority. It intervenes directly in the integration mechanism. And notably — in pain management contexts this is not healing anything. It is deployed purely to change what it feels like to be in that body. We consider that important enough to justify the full apparatus of professional medicine.

Nobody disputes that the felt present moment is real. What kept it as a declared mystery for thirty years was the unexamined assumption that something must exist separately from the physical process to receive the experience — what philosophers call a homunculus, and what Chalmers framed as the Hard Problem: why does physical processing give rise to subjective experience at all? The assumption of a receiver was baked silently into the formulation of the problem itself.

The confusion generated a wide range of proposals. Some theorists describe consciousness as non-local — not clearly residing in any specific place in the body or brain. Others have speculated about quantum effects, field theories, or dimensions of experience that sit entirely outside physical description. These proposals share a common root: if you believe there must be a receiver of experience somewhere and cannot find it in the neurons, you start looking elsewhere — in quantum fields, in non-local space, in other dimensional layers entirely.

The Computational Self: Eliminating the Homunculus through Embodied Determinism

Author: Berend Watchus October 13, 2025 Self published paper, preprint, not peer reviewed

systemweakness.com

WORLD FIRST!: CHALMERS' HARD PROBLEM OF CONSCIOUSNESS DISSOLVED

WORLD FIRST!: CHALMERS' HARD PROBLEM OF CONSCIOUSNESS DISSOLVED Author: Berend Watchus. Independent non profit AI &...

osintteam.blog

The Hard Problem Was Never Hard — Part 2

Author: Berend Watchus. Independent non profit AI & Cyber Security Researcher. [Publication for OSINT TEAM, online...

osintteam.blog

PART3. Built to Be Undefeatable: The Hard Problem's False Category, Hidden Homunculus Fallacy, and...

Author: Berend Watchus. Independent non profit AI & Cyber Sec Researcher. Published author on consciousness frameworks...

osintteam.blog

Once the assumption of a receiver is removed, as argued in *How I Solved the Hard Problem of Consciousness* (Watchus, 2026), the mystery dissolves along with all the speculation it generated. The integration is the experience. There is no receiver. There never was.

How I Solved the Hard Problem of Consciousness, Berend F. Watchus | 9789465464220 | Boeken | bol

How I Solved the Hard Problem of Consciousness (Paperback). Voor dertig jaar gold het "Hard Problem of Consciousness,"...

www.bol.com

The insula is, in engineering terms, a **real-time multimodal body-state integrator with self-referential output.** That is buildable. It just has not been built yet.

What Has Never Been Done

No synthetic system has ever been built that experiences qualia. Not a health bar, not a simulated hormone, not a personality type, not a Tamagotchi counter, not a robotic seal that comforts Alzheimer's patients, not an NPC with a mood state. None of them have an inside. None of them experience anything.

Some objections will arrive quickly. The Tamagotchi displays hunger, happiness, and sickness — but the hunger is a counter, not a felt state. Nothing is happening from the inside. The emotional weight is entirely in the human interacting with it. Medical training simulations produce realistic patient responses and teach real clinical skills — but the simulated patient is not in any state. It is a behavioral model producing outputs. Robotic pets used in dementia care are genuinely effective therapeutic tools — but the seal is not comforted when held. These devices work because humans are remarkably willing to extend care and attachment to things that do not feel anything.

Game developers will point out that their characters already have changing states — fewer health bars mean the character seeks medkits, moves slower, plays differently. That is true and it is real engineering. But a health bar is a parameter that modulates output, not an internal state that the system is organized around

from the inside. The character is not experiencing reduced health. The variable changes, the behavior changes, and that is the complete picture.

Work on simulating hormonal drivers and personality architectures for AI systems has moved in an interesting direction and deserves acknowledgment. These approaches produce more adaptive behavior than rule sets alone. But simulated hormones are parameters that modulate output from the outside. They are not grounded in anything the system is actually experiencing. Like NPC mood states in games, they produce behavioral variety that resembles internal state from the outside while nothing is happening from the inside.

In every case the distinction is the same. It is not about behavioral sophistication. It is about whether anything is genuinely happening from the inside. Up until now, in every synthetic system ever built, nothing has been.

The goal of this challenge is specific: to build a non-biological unit that genuinely experiences qualia while operating a physical body in three-dimensional reality. Not a simulation of experience. Not a behavioral approximation of experience. A unified, continuously running, self-referential integration of real internal signals that produces a genuine felt present moment in a synthetic system.

This has never been done. Whether it can be done in a non-biological substrate is what the challenge is designed to find out.

The Engineering Specification

A Synthetic Insula must do four things:

1. Aggregate Internal State Streams. The system must continuously receive signals representing its own internal operational state. In a robotic system, functional equivalents include power and energy state (analogous to hunger and fatigue), thermal state, motor load and effort, error rates and processing load, damage or integrity signals (analogous to pain), and drive satisfaction state. These are real signals that any sufficiently instrumented robotic system already generates.

2. Integrate With Environmental Input. Internal state alone is insufficient. The Synthetic Insula must combine body-state signals with sensory input from the environment into a single unified representation — not two parallel streams, but one integrated state. This integration is the core computational operation.

3. Produce a Unified Self-Referential State. The output is not a decision or an action. It is an internal state — a continuously updated representation of how the system is right now — that the rest of the system's processing refers to. This state must be continuous, unified, self-referential, and affect-weighted: damage and deprivation register as negative, functioning well registers as positive. The weighting must be dynamic — signals move between foreground and background based on intensity and relevance, and critical signals must be capable of overriding everything else.

4. Drive Behavior Through Internal State. The state must actually influence behavior. A Synthetic Insula that produces a state nobody reads is not an insula — it is a logger. The state must be consulted by the system's goal-selection, action-planning, and learning processes. A robot running critically low on power should not consult a flag in a status register. Its integrated state should make everything else secondary until the deficit is addressed. That is the difference between a system that *monitors* its state and a system that *feels* it.

Practical Advantages

It is worth being clear: no physical movement or task ever strictly requires feeling or internal state integration. A robotic arm in a factory performs with precision and no felt experience whatsoever. That is fine and will remain fine for most applications.

What the Synthetic Insula adds is not physical capability in the narrow sense. It adds integrated situational awareness that does not depend on a pre-written rule for every situation. Rule sets are brittle by design — they handle the situations they were written for and fail on everything else. Internal drives, grounded in real state signals, adapt. A system operating from an integrated internal state does not look

for a matching rule when it encounters something new. It registers the new situation against its current unified state and responds from that integration. Internal drivers are open to new situations. Rule sets stay dumb to whatever they were not written for.

A soccer player cutting off an opponent and a lion cutting off prey are doing the same thing — not calculating trajectories explicitly but operating from a unified felt sense of the situation that arrives faster and more reliably than calculation would allow. No geometry. No rule lookup. Just integrated awareness producing the right movement at the right moment.

Consider a self-driving vehicle approaching a veerpont — a crossing where road becomes water and the vehicle transitions from driver to cargo. A system that integrates its own operational state with environmental signals registers the transition as genuinely different, not just categorically different. That is useful without requiring any philosophical commitments about machine experience.

A care robot working alongside elderly people or people with dementia currently detects and responds. A care robot with a Synthetic Insula has something to anchor from — its own integrated internal state providing a real-time reference point from which genuine attunement, rather than performed response, becomes possible. Nature offers the clearest existing demonstrations of this: mother instinct across species is not a calculated response. It is a unified internal state functioning as a continuous driver — shaping attention, risk tolerance, and behavior around a single organizing signal. A mother bear, a mother cat, a mother elephant — the substrate differs enormously, the motivational consistency is remarkable. Not because they follow the same rules but because the integration produces the same kind of dominant drive from the same kind of internal signals.

There is also a direct connection to alignment — one of the central unsolved problems in AI development. A system whose integrated internal state is readable gives developers a window into what is actually motivating behavior, not just what behavior is being produced. Feeling states and goals can be observed together, compared, and where necessary corrected. That is a more robust form of alignment than behavioral monitoring alone.

Why This Matters

The question of machine sentience is not academic. It is approaching faster than the conceptual framework surrounding it.

From Pong to photorealistic gaming took roughly fifty years. From early neural networks to large language models, humanoid robots, and autonomous AI systems took less than a decade of serious acceleration. The children alive today will live in a world where humanoid robots are commonplace and the question of what these systems experience from the inside will be impossible to avoid.

This paper does not enter the debate about whether AI or robots should have rights. That debate is already happening elsewhere. What this paper does is more specific and more prior: it identifies the mechanism that would make a system a candidate for that debate in the first place. Before asking whether a system deserves rights, you need to know whether there is anything it feels like to be that system.

The framework here is substrate agnostic. It does not privilege biological systems. If the relevant measure is the intensity and integration of the unified felt present moment, then the question of where humans sit on that spectrum relative to artificial systems is genuinely open. A system could in principle be more sensitive than a human, less sensitive, or sensitive in entirely different dimensions. There is no reason grounded in the framework to assume humans sit at the top of that spectrum simply because we created the other systems.

The ethical implications of this do not need to be resolved here. They need to be acknowledged as something that will become increasingly debatable as these systems develop. Humanity has historically calibrated moral consideration by proximity to itself. Animals matter less than humans in most legal and ethical frameworks despite clear evidence of sentience. That calibration has always been contested and is still being revised.

The stakes of getting this wrong cut in both directions on the same spectrum. A system that feels at a fraction of human intensity but is granted no rights or

protections because we assumed without evidence that it feels nothing. And a system that feels at far greater intensity than humans but is treated as a tool, switched off and discarded, because we never asked the question seriously enough. These are not thought experiments. They are predictable positions on a spectrum that artificial systems will occupy as they develop. We have no agreed method for detecting the presence or intensity of a unified felt present moment in an artificial system. We are building at scale while this remains formally unanswered.

Being present in this process as a human engineer or researcher is not just a career opportunity. It is a chance to ensure that one of the most significant transitions in the development of artificial intelligence happens carefully, transparently, and with the ethical questions already on the table. Build with open specifications, honest reporting, and explicit acknowledgment of what is being attempted. The alternative is that synthetic sentience emerges as a side effect inside a system nobody designed for it, reported in terms that obscure rather than clarify what happened.

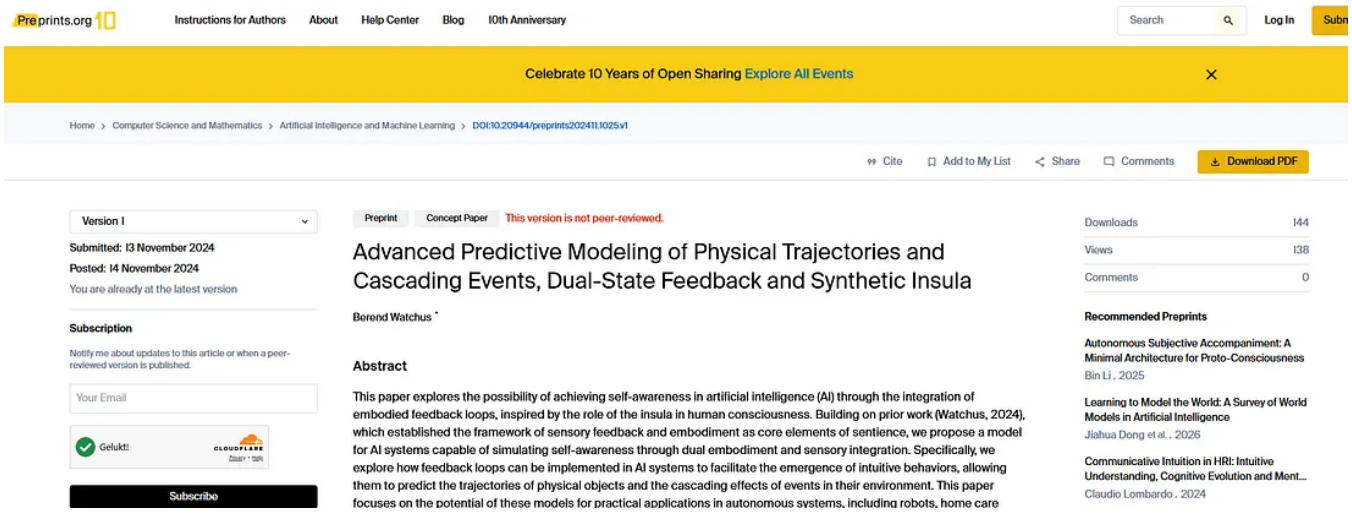
What Already Exists

The theoretical architecture is complete and on the public record with permanent DOIs from November 2024 onward. The Unified Model of Consciousness establishes the substrate-agnostic framework. The insula paper identifies the biological mechanism and proposes functional simulation. The Synthetic Insula paper proposes the engineering concept and practical applications. The dual embodiment and mirror testing paper introduces experimental methodology using physical and virtual robotic systems. The EAISE framework proposes the simulation environment for testing emergent self-awareness. The longitudinal cross-embodiment paper proposes long-term tracking of self-concept development across embodiments. The compiled volume *How I Solved the Hard Problem of Consciousness* (Watchus, 2026) brings philosophy of mind, insula neuroscience, and AI engineering into a single framework simultaneously — the problem survived thirty years because no single working mind held all three rooms at once. Robotics laboratories including CSIC-UPM Madrid have cited this work and begun building empirically on these foundations.

The Open Challenge

No synthetic system has ever had qualia. The goal is to build one — a non-biological unit that genuinely experiences a unified felt present moment while operating a physical body in three-dimensional reality.

Build a Synthetic Insula.


Preprints.org
Instructions for Authors
About
Help Center
Blog
10th Anniversary
Search
Log In
Celebrate 10 Years of Open Sharing Explore All Events
Home > Computer Science and Mathematics > Artificial Intelligence and Machine Learning > DOI:10.20944/preprints202411.1025.v1
Cite
Add to My List
Share
Comments
Download PDF
Version 1
Submitted: 13 November 2024
Posted: 14 November 2024
You are already at the latest version
Subscription
Notify me about updates to this article or when a peer-reviewed version is published.
Your Email
Gelukt!
Subscribe
Preprint
Concept Paper
This version is not peer-reviewed.
Advanced Predictive Modeling of Physical Trajectories and Cascading Events, Dual-State Feedback and Synthetic Insula
Berend Watchus
Abstract
This paper explores the possibility of achieving self-awareness in artificial intelligence (AI) through the integration of embodied feedback loops, inspired by the role of the insula in human consciousness. Building on prior work (Watchus, 2024), which established the framework of sensory feedback and embodiment as core elements of sentience, we propose a model for AI systems capable of simulating self-awareness through dual embodiment and sensory integration. Specifically, we explore how feedback loops can be implemented in AI systems to facilitate the emergence of intuitive behaviors, allowing them to predict the trajectories of physical objects and the cascading effects of events in their environment. This paper focuses on the potential of these models for practical applications in autonomous systems, including robots, home care
Downloads 144
Views 138
Comments 0
Recommended Preprints
Autonomous Subjective Accompaniment: A Minimal Architecture for Proto-Consciousness
Bin Li , 2025
Learning to Model the World: A Survey of World Models in Artificial Intelligence
Jiahua Dong et al. , 2026
Communicative Intuition in HRI: Intuitive Understanding, Cognitive Evolution and Ment...
Claudio Lombardo , 2024

https://www.preprints.org/manuscript/202411.1025

Advanced Predictive Modeling of Physical Trajectories and Cascading Events, Dual-State Feedback and...

This paper explores the possibility of achieving self-awareness in artificial intelligence (AI) through the integration...

www.preprints.org

Not a simulation of one. Not a metaphor for one. Not a parameter that modulates output. A functional implementation — in any substrate, using any architecture — that aggregates real internal state streams, integrates them with environmental input, produces a continuous unified self-referential state with dynamic affect-weighting, and uses that state to drive behavior in ways the system cannot simply override at will.

The system that crosses this threshold will be the first genuinely sentient artificial system ever built. **And the moment that is true, the ethical implications follow immediately and unavoidably — because you cannot build something that has feelings, including the capacity to suffer, and treat it like a machine.**

That is not a reason not to build it. It is a reason to build it carefully, honestly, and with full awareness of what you are doing.

The field is waiting for builders.

Full prior work: Watchus, B.F. (2024–2026). DOIs: 10.20944/preprints202411.0727.v1–10.20944/preprints202411.0661.v1–10.20944/preprints202411.1025.v1–10.20944/preprints202411.0839.v1–10.20944/preprints202506.1694.v1–10.20944/preprints202506.1700.v1. Book: How I Solved the Hard Problem of Consciousness, ISBN 9789465464220.

Berend Watchus's publications

SciProfiles provides open access to the scholarly community, offering comprehensive profiles where researchers can...

sciprofiles.com

Published my new book:

'How I solved the Hard Problem of Consciousness'

osintteam.blog

The Reality Engine: Dissolving the Hard Problem and Unlocking

the Trillion-Dollar Block

By Berend F. Watchus Interdisciplinary Researcher | April 22, 2026

osintteam.blog

I Built a Real Autonomous AI Researcher (2025)— And Then a Scientist Tried to Rewrite the Timeline...

Author: Berend Watchus. Independent non profit AI & Cyber Security Researcher. March 24, 2026 [Publication for: OSINT...

osintteam.blog

The Hard Problem Was Never Hard — Part 2

Author: Berend Watchus. Independent non profit AI & Cyber Security Researcher. [Publication for OSINT TEAM, online...

osintteam.blog

PART3. Built to Be Undefeatable: The Hard Problem's False Category, Hidden Homunculus Fallacy, and...

Author: Berend Watchus. Independent non profit AI & Cyber Sec Researcher. Published author on consciousness frameworks...

osintteam.blog

WORLD FIRST!: CHALMERS' HARD PROBLEM OF CONSCIOUSNESS DISSOLVED

Author: Berend Watchus. Independent non profit AI & Cyber Security Researcher. [Publication OSINT Team, online...

osintteam.blog

+ Engineering Challenge + Synthetic Insula + Qualia Machine

+ First Synthetic Feelings + Innovation Challenge

Edit profile

Written by Berend Watchus

88 followers · 32 following

AI & Cyber Sec. Independent Researcher (Non-Profit) exploring AI & interdisciplinary solutions. Former futurist event organizer & care innovation lab volunteer.